DOCUMENT RESUME

ED 429 336 EA 029 733

AUTHOR	Fuhrman, Susan H.
TITLE	The New Accountability. CPRE Policy Brief.
INSTITUTION	Consortium for Policy Research in Education, Philadelphia, PA.
SPONS AGENCY	National Inst. on Educational Governance, Finance, Policymaking, and Management (ED/OERI), Washington, DC.
REPORT NO	CPRE-RB-27
PUB DATE	1999-01-00
NOTE	14p.
CONTRACT	R308A60003
AVAILABLE FROM	CPRE Publications, 3440 Market Street, Suite 560, Philadelphia, PA 19104-3325 (single copy, free); Web site: http://www.upenn.edu/gse/cpre/; e-mail: cpre@gse.upenn.edu
PUB TYPE	Guides - Non-Classroom (055)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Standards; *Accountability; Educational Administration; Educational Change; Elementary Secondary Education; Performance; Quality Control; School Effectiveness

ABSTRACT

This policy brief reviews developments in standards-based reform and explores new ways for holding schools and districts accountable for discharging their missions. It draws on several Consortium for Policy Research in Education (CPRE) studies of accountability, many of which are still in progress. The report examines characteristics of new accountability systems and focuses on the following: performance, schools as the unit of improvement, continuous improvement strategies, inspections, more accountability categories, public reporting, and consequences attached to performance levels. It discusses issues in designing and implementing new accountability systems, asking such questions as "How is performance measured?" and "What is satisfactory progress?" The brief also describes ways in which to use fair measures in holding schools accountable, such as judging students' satisfactory progress rather than their absolute performance. It offers details on issues such as perverse incentives, complex indices of progress, what to do with students who fall in the middle of the distribution, states' capacity to remedy problems, stability and credibility of programs, public understanding, and student incentives. The report looks at ways in which new accountability systems work, and asserts that when such systems are well designed they can result in improved student achievement and more highly motivated teachers. (Contains 19 references.) (RJM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

CPRE Policy Briefs

Reporting on issues in education reform

Graduate School of Education • *University of Pennsylvania* • *RB-27-January 1999*

The New Accountability

by Susan H. Fuhrman

ED 429 336

EA 027 133

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CPRE Policy Briefs

Reporting on issues in education reform

Graduate School of Education • *University of Pennsylvania* • *RB-27-January 1999*

The New Accountability

by Susan H. Fuhrman

As part of standards-based reform, states and districts are designing new approaches to holding schools and districts accountable for discharging their missions. Virtually every state and thousands of districts are working on developing standards for student learning and aligning student assessments to those expectations. Most are taking the next step which is to use achievement of the standards as a basis for accountability.

The new accountability approaches emerging from this work differ from more traditional systems with respect to one or more of seven factors[1]. District/school approval is being linked to student performance rather than compliance to regulations; accountability is focusing more on schools as the unit of improvement; continuous improvement strategies involving school-level planning around specific performance targets are being adopted; new approaches to classroom inspection are being developed; more categories or levels of accreditation are being developed; school-level test scores are being publicly reported; and more consequences are being attached to performance levels. This policy brief reviews these developments and discusses issues arising from their design and implementation. It draws on several CPRE studies of accountability, most of which are still in progress. Therefore a number of the findings we cite, particularly about the effects of emerging accountability systems, are tentative. However, since policymakers are actively designing and modifying accountability policies, we believe that conveying currently available information about how they seem to be working, even if that information is preliminary, is worthwhile.

Characteristics of New Accountability Systems

A Focus on Performance. States are trying to put more emphasis on student performance and less on compliance with regulations in their accreditation or certification of districts and schools. Traditionally, states have monitored compliance to input and process standards through self-reports and periodic visits. For example, state agencies have monitored pupil/teacher ratios through local forms specifying personnel assignments and student enrollments, and crosswalked them with state certification data to monitor assignment of teachers by field. Minimal curricular requirements have been monitored through on-site inspection of syllabi and other paper work, such as school board minutes, indicating adoption of specific curricula.

More recently, the focus on student performance has led to the monitoring of outcome data such as test scores and graduation rates. Outcome measures (either the level of performance or growth in performance) are now being used as criteria for accreditation. Not only are policymakers interested in progress in the three R's, but they are also including measures of performance in all core subjects in accountability systems.

At the same time, policymakers in many states say they are trying to reduce regulations and compliance monitoring. The idea is to free schools from the old compliance mentality and

EA 029 733

CONSORTIUM FOR POLICY RESEARCH IN EDUCATION

University of Pennsylvania • Harvard University • Stanford University
University of Michigan • University of Wisconsin-Madison

Supported by the National Institute on Educational Governance, Finance, Policymaking and Management, Office of Educational Research and Improvement, U.S. Department of Education

to provide more flexibility so they can maximize student performance using new standards.

Schools as the Unit of Improvement. Whereas traditional accountability systems held districts responsible and focused on district-level activities such as employing teachers appropriately and developing policies to implement state directives, new systems focus on schools. Now measures of success are collected and reported school by school. Districts might also be held responsible for aggregate performance, as in Kentucky, but the important point is that data are now analyzed and reported by school. In the new systems, the state might act—impose a consequence—on these school data, regardless of how good the district-level data look or how the district defines its own responsibilities *vis a vis* school performance.

Continuous Improvement Strategies. Most of the new accountability approaches rest on state-determined performance standards or benchmarks of adequate progress. However, one new approach involves school-site determination of and planning around specific performance targets, such as improved test scores in reading and math. Progress toward meeting locally set targets is the primary measure on which accreditation or state approval hinges. In Kansas, schools develop improvement plans with the assistance of a visiting team of peers within the first 18 months of a 5-year cycle. At the end of the cycle, the same (external) visiting team of 3-4 peers returns to see whether the goals stated in the school's plan have been achieved.

Inspections. Some states are trying to focus their school and district visits on teaching and learning. Instead of paper reviews and central office visits, the new forms of inspection consist of lengthy peer visits that include classroom observation and involve feedback and extensive discussions about practice. In addition to Kansas, just referred to above, Rhode Island is among the states using in-depth reviews. These new inspection approaches are sometimes likened to private agency accreditation, but the aims of the newer state approaches are more ambitious with respect to providing for reflection on practice and attention to student work.

More Accountability Categories. Categories of performance are being expanded to discriminate more discretely among districts and schools. Districts and schools no longer simply pass or fail. Rather, they earn ratings that vary along a continuum. For example, Mississippi has five levels. Those scoring high (levels 4 and 5) are judged to have high-quality programs offering the best opportunity to students. They are allowed some freedom from regulatory policies like teacher certification, time and scheduling, and designing things like K-4 groupings. At the other end are level 1 districts, which are placed on probation.

Public Reporting. More states are publicly reporting district and/or school test scores along with other outcome measures such as attendance and dropout rates. According to the Council of Chief State School Officers (CCSSO), 47 states publicly report scores; 39 of them report school-level data. (CCSSO, 1998)

Consequences Attached to Performance Levels. Increasingly, consequences are being attached to various accreditation levels and/or levels of performance. In addition to public reporting, which draws citizens into the process of evaluating progress and generates its own consequences for schools, some states also provide monetary rewards or other forms of tangible recognition, such as flags or pennants for schools meeting or exceeding performance targets. Currently 14 states provide rewards, mostly monetary in nature (ECS, 1997).

About CPRE Policy Briefs

CPRE Policy Briefs are published by the Consortium for Policy Research in Education (CPRE). The research reported in this issue was supported by a grant (No. OERI-R308A60003) to CPRE from the National Institute on Educational Governance, Finance, Policymaking and Management, Office of Educational Research and Improvement, U.S. Department of Education. *The views expressed in this publication are those of its author and are not necessarily shared by the U.S. Department of Education, CPRE, or its institutional members.*

Sanctions now being proposed for poor performance reach beyond the punishments that states have previously employed for non-compliance. The ultimate punishment for non-compliance in the past has been loss of state aid, a step rarely taken. In the 1980s, six states developed programs for state intervention in severely troubled school districts, sometimes referred to as "academically bankrupt" districts (OERI, 1988). Additiona! states have adopted such approaches in the 1990s; now 31 have sanction provisions in effect. Generally, a series of interventions such as visits by technical assistance teams and the appointment of on-site monitors or conservators precede the ultimate sanctions: state takeover, removal of local governance, or school closure or reconstitution with new faculty and perhaps reassigned students as well. For example, low performers in Oklahoma must develop improvement plans and can ultimately face reassignment of faculty and financial penalty. In Maryland, sanctions can include reconstituting schools and bringing in new staff.

Issues in Designing and Implementing New Accountability Systems

Despite widespread interest in these new approaches to accountability, they are having a hard time getting off the ground. It appears that states and districts are encountering problems putting the new performance-based systems into effect. Studies of accountability systems in 10 states reveal a number of difficult issues that face policymakers and educators (Elmore, Abelmann and Fuhrman, 1996; Massell, 1998). Some of these concerns are technical[2], some have to do with the capacity needed to implement new designs, and some are political.

How is Performance Measured? Measuring student performance involves choosing a set of indicators and instruments. Most policymakers are mindful of the desirability of using multiple measures and not just relying on student test scores (e.g., see Oakes, 1986). So, in addition to student achievement, they include indicators like graduation rates, post-secondary behavior, and attendance.

Each of these indicators involves making decisions about the most appropriate data, but the most difficult and contentious choices have concerned student assessment instruments. For example, when a number of states began to design new accountability systems in the early 1990s, many policymakers hoped to include performance measures using newer types of assessment formats. Performance tasks and open response formats were thought to provide better assessments of the knowledge and skills called for in new standards. However, questions about the reliability and validity of some innovative assessment instruments such as those used in Kentucky and Vermont led to hesitancy about building new accountability systems, especially those with significant rewards and sanctions, around them (Elmore, Abelmann, and Fuhrman, 1996; Koretz, Deibert, and Stecher, 1994; Catterall, Mehrens, Ryan, Flores, and Rubin, 1998; Kentucky Institute for Education Research, 1995). Some early adopters of new forms put multiple choice formats back into their assessment systems, and by mid-decade, most were concerned about "balance," about mixing some open-ended formats with more traditional item formats (Massell, Kirst, and Hoppe, 1997). Limiting open-ended formats also reduces the cost and time requirements of assessments. This is important if states test each individual student, as many parents prefer, and not just samples of students.

States also have to decide whether to develop their own assessments, designed specifically to address their standards, or to purchase commercial assessments, many of which are advertised as being based on national curricular standards. These decisions have become more complex as commercial publishers have included open-response formats and performance tasks in their tests, seeking the balance that states have tried to achieve in their own custom-designed assessments.

Another important issue in measuring student performance is the determination of the standards of performance. Policymakers must not only pick a measure, but they must decide what is good or proficient performance on that measure. It is important that these standards of performance not seem arbitrary to the public. When newer and less familiar assessments are used, it may not be easy to explain to the public why one level of performance is proficient and another is not. This is especially the case if more familiar measures (e.g., SATs or ACTs) seem to provide divergent information, as they did in some Kentucky schools, for example. Avoiding a perception of arbitrariness requires not only clear explanations of the standards-setting process, but also provision of examples of items and performance at each level, so people can see for themselves how performance levels differ and how they reflect different levels of knowledge and skills.

Difficult Issues Facing Policymakers and Educators

- How is Performance Measured?
- What is Satisfactory Progress?
- Holding Schools Accountable for Value-Added
- Perverse Incentives
- Complexity and Trade-Offs with Fairness
- What About the Middle of the Distribution?
- State Capacity for Remedy
- Stability and Credibility
- Process Regulations Don't Go Away
- Public Understanding
- Student Incentives

What is Satisfactory Progress? Many new state accountability systems hold schools responsible for growth in student performance. But how much growth is reasonable to expect in each measurement period? Is it reasonable to expect that the same rate of growth can be sustained over a long period of time? And should the same amount of growth be expected for schools that start at different levels of performance?

Early versions of these new systems used growth projections developed through consultative processes, in the absence of actual evidence about how much growth would be reasonable to expect in each measurement period. In Kentucky, for example, each school is expected to progress to the point where the average student is performing at a proficiency level on the state assessment in all core subjects at the end of a 20-year period; every two years a school must move a tenth of the way toward that. However, now that these systems and their assessment mechanisms have been in place for a few years, some policymakers are trying to set growth curves for schools based on empirical evidence of what is feasible. Policymakers should consult test publishers as well as other states and districts that have used the measures they are considering in order to learn about the average growth paths and variation in performance of schools starting at different levels. Drawing on these experiences, policymakers can determine what growth expectations to build into their new accountability systems.

Holding Schools Accountable for Value-Added. Many policymakers believe that it is unfair to hold schools accountable for factors beyond their control, such as variations in preparation associated with the socio-economic background of students. So they focus accountability systems on satisfactory progress, rather than on absolute performance—all the more reason for setting progress goals carefully. But focusing on growth, often called the "value-added" by the school, generates new controversies. For one thing, many accountability systems compute growth by looking at changes between this year's 4th grade scores and last year's 4th grade scores. In other words, they measure achievement by looking at the difference in performance of two different student "cohorts." But teachers know that classes vary in unpredictable ways. A better measure of "value-added" would track the performance of individual students' scores over time from grade to grade, but that can be very expensive, requiring yearly testing of all students. If the student population is highly mobile, it is even more expensive. In addition, some people say that accepting modest increments of growth in schools starting out with a very low performance base might never get their students to the level where they can participate fairly in society. Rewarding schools for making such gains without motivating them to reach standards of proficiency could be setting lowered expectations for their students. Many states try to incorporate both growth and absolute measures in their systems for this reason.

Measuring learning gains that can be attributed to individual schools is complicated immensely by student mobility. If student turnover is high, it is hard to hold schools accountable for growth as the students being tested change frequently. Schools showing low growth rates overall sometimes show much higher rates of growth for students who have been attending for two or three years. Policymakers can use statistical adjustments to deal with this problem although it is unlikely that they can address it to everyone's satisfaction.

Perverse Incentives. Most educators know that they can bring school-level average scores up by working extra hard with students in a few areas of the assessment, or by focusing their efforts on those students poised to make the largest gains. To avoid such manipulations of the accountability system, which ostensibly is intended to improve the overall performance of all students, policymakers often try to build in safeguards. They monitor student exclusions to make sure that low-performers are not kept out of the testing pool and take security precautions so that no one can tailor test preparation to a few items. In addition, they can require that the scores of the most low-achieving students must improve as well as the average scores of schools. Texas holds schools accountable for showing significant growth in the scores of sub-populations of students, so that educators will focus their efforts on each ethnic group. Similar procedures can be used to ensure that gender differences are not neglected.

Complexity and Trade-Offs with Fairness. In order to use multiple measures, focus on value-added, adjust for mobility and avoid perverse incentives, many states have developed very complex indices to judge school performance and progress. Such formulas may be necessary to fairly account for all the factors influencing performance, but they are difficult to explain to the public. Frequently policymakers themselves do not understand them, and when various constituencies object to sanctions or complain about failure of a school to get a reward, they are unable to defend the system. Policymakers should hold themselves responsible for whatever accountability system they adopt and for making it comprehensible to the larger public. They should balance complexity in the name of fairness with clarity in the name of legitimacy.

What About the Middle of the Distribution? Most of these emerging systems focus attention on the lowest-performing schools, which are subject to sanctions, and the highest-performing schools, which are eligible for rewards. They do little for those in the middle. Schools plodding along at a pace not sufficient to get a reward keep their noses out of trouble, but they are unlikely to get technical assistance (or at least none directly associated with the accountability system) that might boost them into reward categories. Schools that are not improving, but not declining, are likely to avoid sanctions in many systems; they are also likely to be ignored. Evidence of a lack of attention to the middle is the flip-flopping of schools between sanctioned status and non-sanctioned status. In some states, schools in trouble get enough attention to get them out of trouble, but then are left alone for a year or two without further help, and often slip right back into a sanction-eligible category.

State Capacity for Remedy. The reason schools in the middle of the distribution are likely to be ignored, and schools at the bottom frequently get insufficient help, is that states have limited capacity. Not only are agencies hard-pressed due to cuts in state funding, but they also often lack the kind of instructional expertise necessary to make a difference in troubled schools. To generalize, states employ a series of remedies for schools judged to be low-performing, ranging from technical assistance to takeover to actually closing the schools, by either redistributing students elsewhere or by reopening it with new staff. But often the remedies are vague and the technical assistance is meager. One part-time person might serve as an advisor, for example. The new federal Comprehensive School Reform Design Program, which provides $50,000 annually for schools employing external designs, might help low-performing Title I schools get assistance they wouldn't otherwise. However, the most troubled schools may be too distressed to pick or profit from a good design. They need an infusion of assistance nonetheless, and few states adequately budget for it. It appears that many states fail to appreciate just how costly the remedy or sanction side of the accountability system can be. They know how to cost out rewards, but tend to underestimate the cost of sanctions.

Stability and Credibility. Political change is threatening to accountability systems that depend on educators' faith that the consequences they are told to expect will really happen. Leadership turnover and change of heart—and the tendency to back away from severe sanctions at the end of the day, which has happened in some places—undermine the credibility of these systems. An illustrative anecdote is the famous "$300" problem in Kentucky. In the 1980s, the legislature passed and then failed to fund a promised bonus for teachers. The new accountability system enacted in 1990 had to overcome this history and convince teachers that they really would get the rewards built into the new system. Skepticism was a major barrier to making the new system work. Legislators might consider putting reward dollars in escrow to reassure teachers that they really will be available in the future.

Those Process Regulations Don't Go Away. Bearing down on performance makes sense if the state simultaneously relaxes from process regulations that might restrict flexibility in reaching the new performance goals. Examples are mandates to teach students or teachers specific topics, like Lyme Disease prevention, that might be worthwhile but that are not related to the state's performance goals. Despite rhetoric about more autonomy in return for greater accountability, however, most states have not repealed much of the existing process regulation. Opposition from those attached to the status quo, comfort with traditional arrangements, and inertia, among other reasons, make it unlikely that policymakers will revisit old policy when they pass new policy. Some regulations protect powerful political constituencies, like special interest groups seeking a niche in the school curriculum. And many regulations are intended to ensure that minimal standards are met in districts of varying wealth, location, size, and capacity. These regulations form the cornerstone of traditional approaches to equity, and whether or not they are effective, it is still difficult to argue for their removal in the face of persistent differences in resources, teacher quality, and program offerings.

Public Understanding. Not only are new accountability systems likely to involve complex formulas, but also they often entail many other unfamiliar elements: standards-based tests given to samples of students at key grades (as opposed to the familiar pattern of norm-referenced tests being given every year and yielding scores for each child relative to other children); new roles for states intervening in schools; and new publicity attached to performance. It is important that the public be educated about all of these aspects; without public support, new systems will wither over time. It is particularly important that people accept the need for students to do well on the measures selected; it should be clear that performance on state assessments affects important life outcomes. Otherwise, parents will not encourage students to work hard in preparation, not understand why state money should go to rewarding schools that do well on such tests, and not support sanctions accorded to schools that don't do well.

Student Incentives. Most of the new systems do not embed consequences for students in their accountability approaches. These systems are not designed to motivate students. Yet education is co-produced; teachers' success is dependent on students' efforts to learn. Many teachers argue that ignoring student consequences is unfair. They can be penalized if the students don't do well but there is nothing to motivate students to take the tests seriously. Gradually, some states are attaching promotion and graduation consequences to their new assessment measures, phasing in the requirements so that schools have a lot of time to give students adequate opportunity to learn the necessary material.

International evidence suggests that curriculum-based examinations at the time of high school completion can spur higher overall achievement. Students from countries using medium- and high-stakes assessment systems outperform students from other countries, at a comparable level of economic development, that lack such systems. In addition, qualifications for entry into secondary school teaching are also higher in nations with curriculum-based external exit examinations. Evidence from Canada shows this same pattern among that nation's provinces. Not only did 13-year-olds from Canadian provinces with exit exams know more science and mathematics than students in other provinces, they watched less TV and talked with their parents more about schoolwork. Schools in provinces with external exams were more likely to employ specialist teachers of mathematics and science; hire math and science teachers who had studied the subject in college; have high-quality science laboratories and experiments; schedule more hours of math and science instruction; and assign more homework in math, in science, and in other subjects. Clearly, Canadian external exams have not lowered the quality of instruction, they appear to have enhanced it (Bishop, 1998).

Even low-stakes curriculum-based exit exams such as the New York State Regents exams appear to have an effect. Holding student demographics constant, New York State students do significantly better than students in other states on the SAT test and the NAEP math assessments (Bishop, 1998). New York is currently phasing in a requirement that all students must pass the Regents exams to graduate, creating a high-stakes accountability environment for students.

How Do New Systems Work?

New accountability systems focus attention on performance, seeking to inspire school personnel to attend more to student achievement. What we don't yet know is the quality and duration of the instructional changes that occur in responding schools. We don't know whether such systems cause teachers and administrators to attend to the kind of capacity-building that will enable them to improve instruction and learning over time.

Getting Attention. Evidence on effects comes from studies of high-stakes systems. Clotfelter and Ladd (1996) found higher pass rates in reading and math in Dallas schools, where a complex reward system is in place, than in other Texas cities, even after adjusting for other reform efforts in those cities. In addition, the disaggregated, sophisticated information collected to operate the incentive program in Dallas provided a useful managerial tool.

CPRE has conducted interviews and large-scale surveys of teachers and principals in two research sites with new accountability systems: North Carolina's Charlotte-Mecklenburg School District and Kentucky (Kelley, Milanowski, and Heneman, 1998). We found that setting student achievement goals for a school helps provide teachers with a focus for their work and increases the energy devoted to instruction. New systems help channel teachers' work to the most important goals of the system, largely those included in the performance measure, which, in the cases we are studying, is student achievement in the core academic subjects.

Teachers placed value on the consequences included in these accountability systems, and those consequences helped motivate them to work in more focused ways to produce improved student achievement. In the Kentucky and Char-

lotte-Mecklenburg studies, we found that teachers positively valued personal satisfaction from increasing student learning, professional recognition for doing a good job, and receiving a monetary bonus. Bonuses seemed to be less of an incentive for many teachers than a "thank you" for a job well done, but teachers tended to agree that the bonus was an important symbol of accountability or efficiency to the public. Negative outcomes, which were equally motivating, included increased pressure and stress to improve results, fear of being labeled as a "school in decline" and the accompanying professional embarrassment, loss of freedom through some state directed assistance or "takeover," and expanded work hours.

Additional findings about the bonus part of these programs were that teachers who reported being most motivated by the program and wished to see it continue were also those who were most dissatisfied with their current level of pay, those who felt it was fair to receive a bonus for improving student achievement, and those who felt that administration of the bonus process was fair.

Our conclusion is that, contrary to the views offered by some commentators, money bonuses are valued by teachers and can be motivating, and sanctions such as school reconstitution or identification as a school in decline are also valued (though negatively) and can also function to motivate. Both get teachers' attention.

Variable Responses. It is also clear from our work, however, that teachers and schools vary in their responses to these accountability systems. Several factors help account for this variation. If schools had previously received a reward, the possibility of another positive outcome tended to be a primary motivator. If a school was in a non-improving mode, the negative outcomes seemed more salient.

Much of the variation in response is related to individual and organizational capacities. Teachers believe that their knowledge and skill and school conditions are the critical elements that help them accomplish accountability goals. Principal leadership skills, the opportunity for feedback on results, curriculum alignment, professional development, and the sense of professional community at the school are more likely to be assessed positively in schools that accomplished their goals for improving student achievement than in schools that did not.

In fact, without strong leadership, a cohesive professional community, and access to good professional development, a school's response may be short-term and superficial. On the whole, the schools operating under severe sanctions (reconstitution and probation) that CPRE is studying in San Francisco and Chicago do not appear to be making fundamental changes in their core processes. These schools seem to be placing considerable emphasis on test preparation. Some may incorporate some structural changes (such as breaking up into small schools) but few appear to be making extensive or deep efforts to rethink their instructional program (O'Day, in press).

An important question for research is how these new state and local accountability systems interact with variations in school capacity. If, as previous research predicts and current studies confirm, factors such as teacher knowledge and skill, principal leadership, professional community, etc. "enable" positive school responses to external accountability systems, how do we work to foster these conditions? In other words, how can accountability systems lead to effective capacity-building? Are certain approaches to accountability more likely than others to engender capacity-building? Over the short-term? Over the long-term? How else can policy provide for capacity-building *in addition to* establishing accountability systems?

Relationship Between Internal and External Accountability Systems. One way to think about the relationship between school capacity and new accountability systems is to ask about pre-existing conceptions of accountability within schools. What are the within-school or "internal" norms about accountability and how do they interact with the "external" systems created by state or district policy? When new external systems, such as the ones we have been describing, hit schools, can we expect variations in responses based on how schools already think about their responsibility for student learning? In other words, "When accountability knocks, will anyone answer?" In 20 public, parochial, independent and charter schools, CPRE found that internal accountability systems varied widely and were not necessarily related to the external accountability system (e.g., whether the school was a charter school). It is important to note that none of these schools were in Kentucky or other settings with strong new accountability systems (Abelmann and Elmore with Even, Kenyon, and Marshall, 1999).[3]

In the typical school in our sample, individual teachers' conceptions of their responsibility exercised the greatest influence over how schools addressed the accountability problem. For the majority of schools we studied, questions about the collective purposes of the enterprise were answered by

the aggregation of individual teachers' decisions. The views of individual teachers about their capacity and that of their students determined what, if anything, was done to improve performance, not collective deliberation or explicit management decisions. In such circumstances, the school's "solution" to the accountability problem—to whom, for what, and how?—became simply a collection of individual, often idiosyncratic, judgments by teachers, growing out of their backgrounds, capacities, and individual theories about what students can do. These judgments were powerfully influenced by teachers' preconceptions about the individual traits of their students, as well as the characteristics of students' families and the communities from which they come, and typically uninformed by systematic knowledge of what these students might be capable of learning under different conditions of teaching.

Hence, in these schools, teachers assigned the most powerful causality in their own conceptions of responsibility to factors over which they, as teachers, have little or no control, and they assigned the least powerful causality to those conceptions over which they have the greatest control, the conditions of teaching and learning in the school. Schools operating in this mode were characterized by an emphasis on order and control (the one collective expectation on which it was possible to get consensus in an essentially isolated work environment) and by low expectations for student learning, usually wrapped in theories about the relative deprivation of students, their families, and their communities.

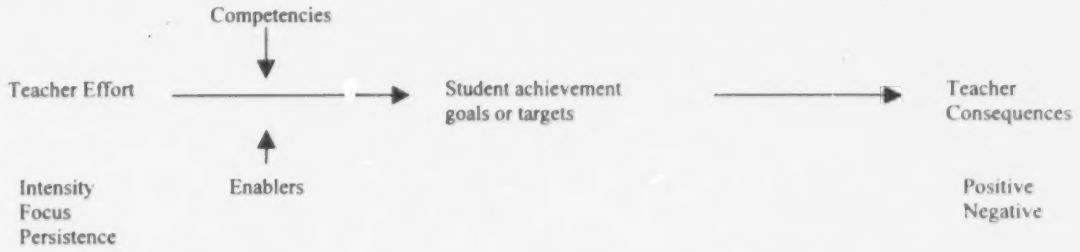
In these schools, teachers and principals often dealt with the demands of formal external accountability structures (curriculum guidance, testing, and the like) either by incorporating them in superficial ways—claiming, for example, that they were consistent with existing practice when they clearly weren't—or by rejecting them as "unrealistic" for the type of students they served. As noted, none of the schools in the 20-school study were located in strong external accountability environments. Some were in districts or states in the early stages of developing stronger external accountability systems. The charter laws operating in these states were in the early stages of implementation, and these states hadn't yet confronted the issue of charter renewal directly. So the observed weak effects of external accountability systems may simply be attributable to the specific policy environments.

But, it seems highly unlikely that schools in which all questions of accountability are essentially questions of individual teacher responsibility will be capable of responding to strong, obtrusive external accountability systems in ways that lead to systematic, deliberate improvement of instruction and student learning. The idea that a school will "improve" its instructional practice, and therefore the overall performance of its students, implies a capacity for collective deliberation and action that schools operating in the typical mode observed in this CPRE study simply did not display. Where virtually all decisions about accountability are decisions taken by individual teachers based on their individual conceptions of what they and their students can do, it seems unlikely that these decisions will mystically produce a strategy for overall improvement for the school. So for schools operating in this mode, the key question for future research on the effects of external accountability systems is whether schools can, or will, respond by developing congruent internal expectations and accountability systems. And, perhaps more important, a related question is how these schools develop the capacity to generate these new internal norms and processes.

Schools that have developed clear internal expectations and/or accountability systems raise a different set of issues about the relationship between internal and external accountability. Our research suggests that such schools answer the question of what they are accountable for in very different ways—some focus on relatively low-level therapeutic purposes, some on higher-level aspirations for student performance. So for these schools, the issue is the degree of alignment between their internal expectations and accountability systems and the demands of external systems, as well as the level of conflict and accommodation that grows out of the confrontation between internal and external accountability. Are schools that manifest some capacity to deal collectively with the internal accountability problem more likely to adapt and align their internal norms and systems to the requirements of external systems, or are they likely to be more resistant to changing their norms and systems? Do such schools have the necessary capacities to do the work of accommodating and adapting new external requirements? The existence of collective internal expectations and accountability structures, in other words, does not necessarily predict exactly how a school will respond to new external requirements. These are questions to pursue in future research.

New Accountability Systems: Getting Attention vs. Getting Capacity. The research that we have just summarized shows that new accountability systems can be motivating, especially in the presence of certain enabling conditions, and that internal, shared beliefs about the importance of student performance are likely to be key enablers.

Figure 1
A Model of Teacher Motivation



Previous research calls attention to both "will" (motivation) and "skill" (capacity) as influences on improving schooling (e.g., McLaughlin, 1987). In our studies of the influence of accountability systems, we have been working with a model of motivation developed from Expectancy Theory (Vroom, 1964) and Goal-Setting Theory (Locke and Latham, 1990) (see Figure 1). The model suggests that a teacher will be motivated to try to reach the school's student achievement goals to the extent that she (a) perceives a high probability that teacher effort will lead to students' reaching achievement goals (expectancy perception); (b) perceives a high probability that goal attainment will lead to certain consequences or outcomes such as a bonus award (instrumentality perception); and (c) places value, either positive or negative, on these outcomes.

What we have learned is that new accountability systems seem to work primarily on (b) and (c); i.e., they relate goals (student achievement) to outcomes or consequences (both positive and negative) that teachers generally value. These are important contributions. As noted previously, teachers have often been skeptical about promises that reaching goals will have consequences; in the systems we have studied, this skepticism was overcome. Teachers believed that the bonuses or the sanctions would actually take place. And, they find both the positive consequences, including bonuses, and the negative consequences motivating.

With respect to (a), however, the expectancy that teacher effort will actually lead to the desired level of student achievement, accountability systems appear to provide only a *part* of the answer. Goal-setting theory (Locke and Latham, 1990) suggests that the nature of the goals themselves can affect the effort-to-goal calculus. The clearer and more understandable the goals, the more likely teachers will think that their effort will lead to reaching them. So, to the extent that new accountability systems include clear and specific goals, teacher expectancy may be affected. This finding certainly argues that attending to several of the design issues mentioned previously—using meaningful measures, setting reasonable expectations of progress, and establishing comprehensible and fair systems—are important. Further, when new systems include technical assistance, such as the help provided by Distinguished Educators assigned to schools in decline or crisis in Kentucky, they are attending directly to factors influencing expectancy: teacher knowledge and skills, curriculum alignment, and other enabling conditions. Reward schools in Kentucky that rebounded from decline status specifically attributed their success to the assistance of a Distinguished Educator (Kelley, 1998).

Apart from the technical assistance provided to schools in trouble (which, as noted previously, is frequently dropped when the school improves), accountability systems as currently designed do not directly work on the capacity issue—on the teacher competencies and school conditions that affect the effort-to-achievement calculus. Their design is based on one of two assumptions: either schools already possess the capacity, but not the will to meet goals, or, once the goals are clear and valued consequences are attached, schools will shop for or find the capacity they need to meet the goals.

However, we have already seen that the mere imposition of a new accountability system—even ones like the Kentucky and Charlotte-Mecklenburg systems where the goals are thought to be clear and outcomes both sure and motivating—does not unleash some hidden capacity. Schools clearly vary in their response to accountability systems, and that variation is strongly associated with existing school and teacher capacities. As to whether accountability systems cause schools

to seek help and find or build the necessary capacity over time, only research over a longer period can tell. We are conducting such research, studying schools operating in established accountability systems as well as in systems just coming on line. We will look at different types of accountability systems, including ones that attend to student incentives. Our work will examine both the demand side—whether schools seek capacity-building help—and the supply side—the adequacy and quality of capacity-building interventions available in the public, private, and non-profit sectors. If accountability systems do not create sufficient demand over time and/or if the existing supply of assistance is unsatisfactory (or unavailable to schools that lack sufficient fiscal resources), then states and districts will have to attend directly to capacity-building themselves if they want to increase system performance. They will not be able to rely on accountability systems alone to leverage significant improvement in student achievement.

Also clear from our research is that capacity-building efforts must attend to internal norms about accountability. Certainly teacher knowledge and skills are key, but so are teacher expectations about student learning. Teacher beliefs about whether their students will benefit from their effort exercise a strong influence on the expectancy calculus. As we have seen from our 20-school study of internal accountability, many schools' teachers have no collective sense of responsibility for learning, and individual teachers are strongly influenced by what they think their students can or cannot do. They have had no experience with students from disadvantaged backgrounds learning challenging material; giving them such experiences and showing them how their children can learn to high standards are important aspects of capacity-building. In addition, the shared norms and processes underlying collective school improvement must be developed.

Conclusion

New accountability systems that are well-designed (with fair, comprehensible, meaningful, and stable features) are associated with improved student achievement when adequate capacity to improve instruction is present in schools or can be provided by an outside partner. Generally, teachers find the new systems motivating. But, in the absence of explicit attention to capacity, the new systems are insufficient approaches to improving student achievement. Our continuing research will provide better information about the role of new accountability systems in encouraging sustained capacity-building and we look forward to reporting the findings in a future policy brief.

About the Author

Susan Fuhrman is both Dean of the Graduate School of Education and the George and Diane Weiss Professor of Education at the University of Pennsylvania, and also serves as Chair of CPRE's management committee. She has conducted research on state education reform, state-local relationships, state differential treatment of districts, federalism in education, incentives and systemic reform, deregulation and legislatures and education policy. A prolific writer, Fuhrman has written a plethora of book chapters, journal articles, and reports. In addition, she has edited or co-edited the volumes, *Designing Coherent Education Policy: Improving the System* (1993), *Rewards and Reform: Creating Educational Incentives that Work* (1996, with Jennifer O'Day), *The Politics of Curriculum and Testing* (1991, with Betty Malen), and *The Governance of Curriculum* (1994, with Richard F. Elmore).

Acknowledgments

The author would like to thank Rolf Blank of the Council of Chief State School Officers for his insightful comments on an early draft of this brief, and Jennifer Dounay and Judie Mathers of the Education Commission of the States for verification of facts presented in this publication. Finally, the author is grateful to the National Institute on Educational Governance, Finance, Policymaking and Management, Office of Educational Research and Improvement, U.S. Department of Education for its continuing support of CPRE.

References

Abelmann, C., and Elmore, R. F., with Even, J., Kenyon, S., and Marshall, J. (1999). *When accountability knocks, will anyone answer?* (CPRE Research Report No. RR-042). Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.

Bishop, J. (1998). *Do curriculum-based external exit exam systems enhance student achievement?* (CPRE Research Report No. RR-040). Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.

Catterall, J. S., Mehrens, W. A., Ryan, J. M., Flores, E. J., and Rubin, P. M. (1998). *Kentucky instructional results information system: A technical review*. Frankfort, KY: Author.

Clotfelter, C., and Ladd, H. (1996). Recognizing and rewarding successful schools. In H. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 23-63). Washington, DC: The Brookings Institution

Council of Chief State School Officers. (1998). *State education accountability reports and indicator reports: Status of reports across the states*. Washington, DC: Author.

Education Commission of the States. (1997). *Education accountability systems in 50 states*. Denver, CO: Author.

Elmore, R. F., Abelmann, C., and Fuhrman, S. H. (1996). The new accountability in state education policy. In H. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 65-98). Washington, DC: The Brookings Institution.

Kelley, C. (1998). The Kentucky school-based performance award program: School-level effects. *Educational Policy, 12*(3), 305-324.

Kelley, C., Milanowski, A., and Heneman, H. G. III. (1998, April). *Changing teacher compensation: Cross-site analysis of the effects of school-based performance award programs*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Kentucky Institute for Education Research. (1995). *An independent evaluation of the Kentucky instructional results information system (KIRIS)*. Frankfort, KY: Author.

Koretz, D., Deibert, E., and Stecher, B. (1994). *The Vermont portfolio assessment program: Findings and implications*. Washington, DC: RAND Institute on Education and Training.

Locke, E. A., and Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice-Hall.

Massell, D. (1998). *State strategies for building capacity in education: Addressing the needs of standards-based reform*. (CPRE Policy Brief No. RB-25). Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.

Massell, D., Kirst, M., and Hoppe, M. (1997). *Persistence and change: Standards-based reform in nine states*. (CPRE Research Report No. RR-037). Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.

McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis, 9*(2), 171-178.

Oakes, J. (1986). *Educational indicators: A guide for policymakers*. (CPRE Occasional Paper No. OPE-01). New Brunswick, NJ: Rutgers University, Center for Policy Research in Education.

O'Day, J. (In press). *Reconstitution as a remedy for school failure*. (CPRE Policy Brief). Philadelphia, PA: University of Pennsylvania, Consortium for Policy Research in Education.

Office of Educational Research and Improvement (1988). *Creating responsible and responsive accountability systems: A report of the OERI state accountability study*. Washington, DC: Author.

Vroom, V. H. (1964). *Work and motivation*. New York: Wiley.

End Notes

1. This discussion includes excerpts from Susan Fuhrman, "Evaluation of Performance in the United States: Changes in Accountability" prepared for the Organization of Economic Cooperation and Development, August 1994.

2. New developments in testing technology, such as internet-based computer-adapted testing, may alleviate some of the technical problems in the future (The Consortium on Renewing Education, November 1998. *20/20 Vision: A Strategy for Doubling America's Academic Achievement by the Year 2020*. Nashville, TN: Peabody Center for Education Policy, Vanderbilt University).

3. The remainder of this section draws heavily on language in Abelmann, et al., 1999.

About CPRE

The Consortium for Policy Research in Education (CPRE) unites five of the nation's leading research institutions in an exciting venture to improve student learning through research on education policy, governance, and finance. Members of CPRE are the University of Pennsylvania, Harvard University, Stanford University, the University of Michigan, and the University of Wisconsin-Madison.

CPRE conducts research on issues such as:

- education reform
- student and teacher standards
- state and local policymaking
- education governance
- school finance
- teacher compensation
- student incentives

To learn more about CPRE or CPRE publications, please call 215/573-0700. CPRE's website address is:

http://www.upenn.edu/gse/cpre/

Nondiscrimination Statement

The University of Pennsylvania values diversity and seeks talented students, faculty, and staff from diverse backgrounds. The University of Pennsylvania does not discriminate on the basis of race, sex, sexual orientation, religion, color, national or ethnic origin, age, disability, or status as a Vietnam Era Veteran or disabled veteran in the administration of educational policies, programs or activities; admissions policies, scholarships or loan awards; athletic, or University administered programs or employment. Questions or complaints regarding this policy should be directed to Executive Director, Office of Affirmative Action, 1133 Blockley Hall, Philadelphia, PA 19104-6021 or (215) 898-6993 (Voice) or (215) 898-7803 (TDD).

Cyberspace Update

Have you had a chance to check out CPRE's website (http://www.upenn.edu/gse/cpre) lately? If not, we encourage you to do so. We have made some major changes including making more publications available on the World Wide Web. For example, all *CPRE Policy and Finance Briefs* can now be downloaded. Detailed research project information is now available. We suggest checking out our Paper of the Month, one of the most popular features on the website. Finally, we would love to get your feedback. What can we do to improve the site and better meet your needs? Let us know your thoughts by emailing our webmaster, Robb Sewell, at *robbs@gse.upenn.edu*.

CPRE Policy Briefs

NON PROFIT
U.S. POSTAGE
PAID
PERMIT NO. 2563
PHILADELPHIA, PA

Graduate School of Education
University of Pennsylvania
3440 Market Street, Suite 560
Philadelphia, PA 19104-3325